4-1-2019

Exploring the Influence of Social Network Geography on Long-Distance Travel Behavior

Jonathan Dowds
University of Vermont, jdowds@uvm.edu

Hannah Ullman
University of Vermont

Sarah Howeter
University of Vermont

Lisa Aultman-Hall
University of Vermont

A Report from the University of Vermont Transportation Research Center

Exploring the Influence of Social Network Geography on Long-Distance Travel Behavior

Final Report

April 2019

Sarah Howerter, Hannah Ullman, Jonathan Dowds, and Lisa Aultman-Hall

Exploring the Influence of Social Network Geography on Long-Distance Travel Behavior

April 2019

Prepared by:

Sarah Howerter
Hannah Ullman
Jonathan Dowds
Lisa Aultman-Hall

Transportation Research Center
Farrell Hall
210 Colchester Avenue
Burlington, VT 05405

www.uvm.edu/cems/trc/research-reports/

Acknowledgements

This project was funded by the National Center for Sustainable Transportation, a national University Transportation Center of the USDOT. We appreciate the many discussions with Dr. Jeffrey LaMondia of Auburn University for this project.

Disclaimer

Table of Contents

List of Tables

List of Figures

Abstract

Over the last two decades, a limited number of studies have sought to measure attributes of one's social network and connect these measures to travel. Increasingly, burdensome social network surveys include a contact's location. This study focuses on long-distance travel, itself a challenge to quantify. The People in Your Life survey was a pilot mail-back questionnaire with 110 respondents in three regions of the United States. A method to characterize social network geography was proposed using not only distance between ego and contacts but also contact to contact distance. The new approach is able to incorporate the geographic extent of the networks when compared to the more basic approaches. Moreover, reasonable clusters were created using this small sample. The results agree with prior studies that social network extent is related to types and levels of travel. The research here was not conducted on a full or comprehensive social network, we only surveyed 13 total contacts, suggesting that there is merit to the idea that representative, but not comprehensive, social networks may be adequate for transportation-related research. If future research could comprehensively validate this proposition, the burden of adding social network measures to travel surveys would be reduced and potentially manageable.

Keywords: long-distance travel, social networks

1. Introduction

Transportation planning professionals began considering the influence of a traveler's social network in the early 2000's beginning with publications by Axhausen (1-4) and his coauthors Larsen and Urry (5-6). Despite these studies, most transportation planning still sees travel as a derived demand stemming from a desire to participate in activities spread across space without regard for the location of a social network. These early works on social networks and travel drew on both social science and transport planning approaches and were largely exploratory, suggesting methods for an analysis of social networks that focused on their relation to or influence on travel. Some studies examined the influence of the rapid changes in communication technologies and were based on assumptions that the effective impact of technology on the geography of costs and locations would impact travel and social patterns profoundly (4). Early studies were also couched in the sense that leisure travel had been growing and pointed to its social element. Significant research beyond the field of transportation has been conducted on social networks, both around data collection and analysis. However, the spatial dimension and location of one's contacts has primarily been left out because of the substantial burden it places on survey respondents. This spatial dimension or geographic embeddedness of a social network is a key element in its influence on travel behavior, an inherently spatial phenomenon.

In this research, we are interested in the relationship between the spatial or geographic attributes of contacts in a traveler's social network and the potential influence it might have specifically on longer distance or intercity travel. Within this context one cannot ignore the assumption and observation that a large portion of leisure travel is continuing to increase across many countries, and this necessitates examination of social networks as a key explanation of a significant portion of long-distance travel demand. There has been a relative lack of data on long-distance travel behavior due to the infrequent nature of these trips and therefore limited inclusion in most daily travel surveys. Long-distance travel data are challenging to collect in recall surveys. In seeking to measure the relationship between social network geography and long-distance travel, one recognizes that both comprehensive, non-biased social network information and long-distance travel are challenging to measure. Within this context, this paper 1) reviews previous work related to both social networks and travel; and 2) investigates a method to characterize social network geography using a pilot survey dataset with a limited number of social contacts; and 3) evaluates how social network type relates to levels of long-distance travel. Our objective is to advance social network data collection for transportation planning and contribute to methods to better characterize the geography of a social network.

The rest of the paper will be organized as follows. A literature review is followed by a description of the People in Your Life Survey (PiYL) administered to 110 individuals in the states of California, Vermont, and Alabama in the United States. In the results section a cluster method is used to classify participants based on their social network geography and preliminary assessments of the relationship to their self-reported long-distance travel frequency is reported.

2. Literature Review: Ego-Centric Social Networks and Travel Behavior

Social structures, resources, and processes have been of interest to social scientists since the late 1950's. They have been studied to observe and analyze organizational structures, mental health, equity, personal choice, social influence, and even the characteristics of social networks themselves. The structures that people are socially embedded in are multifarious and complex and there are multiple ways to measure and describe them. The network approach, presented by Wellman (7), studies social structure by analyzing the patterns of ties linking the members of the network. These complete social networks may be defined in full, gathering all ties linking all of the people in the closed full population (8). The most common method in transportation-related research is to only gather certain sets of social network ties that may be of importance in the generation of travel.

For the purpose of predicting behaviors or choices of individuals, "egocentric" social networks are most often used. An egocentric social network is one that consists of a set of ties, contacts, or "alters" surrounding a sampled individual, or "ego." In 1984, Burt suggested the standard method for collecting these types of networks was the General Social Survey (GSS) (9). He utilizes survey questions referred to as name generators (10), which elicit the names of persons of some specified relation to the sampled "ego" individual, such as persons with whom the respondent has discussed personal matters during the past 6 months. These names are then used for measurement in the remainder of the survey instrument.

Name generators have been used in most transportation-related social network surveys and have been studied in-depth by some (11, 12). For long-distance travel behavior, the set of contacts that researchers could strive to capture might not be the same as those sampled for daily activity-travel. The pilot survey developed by our group reflects an assumption that any social contacts living at a distance that might either induce trips for relationship maintenance or the opportunity to visit a destination with social ties are of importance. However, for long-distance social networks there are additional challenges such as the fact that relationships at a distance may be less strong and thus less easily reported, and that respondents may recall contacts in a geographically associative manner, filling out their generated list with people in one locale without recalling members from diverse other locations. Methods such as the multiple generator random interpreter (MGRI), suggested by Marin (12), which include of a full set of name generators with interpreters only administered to a random subset of the contacts listed help to minimize associative biases and might prove to be of use for collecting long-distance social networks. But here we seek to evaluate the utility of a very small and limited contact set for assessment of geographic extent.

Egocentric social network data have attributes at three different levels: the ego level (socio-demographics and characteristics of the respondent), the ego-network level (aggregate features of the personal network such as size, total interaction frequency, and homophily; referring to how similar the individuals in the network are to the ego, in terms of sociodemographic variables), and the ego-contact level (interpersonal characteristics between the ego and each contact such as tie strength, contact frequency, geographic distance) (13).

From these variables different properties of the network can be analyzed such as spatial distribution of contact locations or the nature of activity-travel behavior. Models can be set up to predict network size, frequency of interaction, social activity participation, or activity-travel decisions.

The earliest work that sought to measure or predict social network size was only concerned with the number of contacts or network members since at that time, locations of the contacts were not being collected (14-15). The desire to relate social networks to travel motivated the collection of these locations, but the quantification of their geography as an ego-level characteristic is not simple or straightforward.

Five somewhat recent collections of social network data from four different countries (Canada, Switzerland, the Netherlands, and Chile) were studied in comparison to one another, considering distance patterns of social contacts with multi-level modeling (16). They considered each ego-contact relationship and geographic great circle distance and aggregated all of the data to compare distance distributions for each dataset. They observed differences in the decay rates of the distributions potentially due to factors "such as the ratio between wage and transport costs, availability of mobility tools and the influence of immigration" (16). They also used two different approaches to model the ego-contact geographic distance as the dependent variable. In order to estimate the model they jointly structured the datasets into three levels: depending on contact characteristics, ego socio-demographic and personal network characteristics, and the study area. They found that the availability of transport and communication relative to income plays a key role in the spatial distribution of contacts (16). Carrasco et al. in their Connected Lives Study collected and mapped complete personal social networks, including connections between contacts within the networks, and found associations between personal (ego) characteristics as well as network composition and ego-alter distance patterns (17).

Frei and Axhausen (18) using the same, aforementioned dataset from Switzerland similarly focused on the spatial dimension of social networks as defined as a link (ego-contact level) attribute and not as a network structure. Two stochastic models were explored for spatially embedded social networks and showed that the observed exponential distribution of tie distance can be explained with a relatively simple homophily model. They also suggested that the great circle distance is probably not the most appropriate spatial measure in the context of transportation research and suggested that some estimation of travel time and cost should be used instead.

If planners intend to utilize measures of social network geography as a predictor of long-distance travel behavior, they must develop a variable or method to characterize it at the ego-network level rather than the ego-contact level. The simplest way to capture this would be to sum the distances of all ties in the network, however, this fails to capture distribution patterns such as clustering or ego isolation (19).

The most common method for the measurement of network spatial "size" is the confidence ellipse method. The confidence ellipse method is a parametric method defined by a fixed percentage confidence region, first presented for the measurement of a person's activity space by Schönfelder (20). This has become the standard measure for egocentric social networks because it is easily computed and has been found to correlate with other more difficult to calculate method (19, 21-24). The area of the ellipse, centered on the ego's home location, represents the network size and the calculation of the ellipse works under the assumption that the locations are normally distributed. The original use of the confidence ellipse was for the measurement of a person's daily activity space (20), which tends to have a smaller localized spatial distribution. Adapting this method for globally distributed, egocentric personal network extent further diminishes the accuracy of this tool since more of the area captured by the ellipse is likely to consist of empty space such as bodies of water or deserts.

Axhausen and Frei (19) utilize the confidence ellipse method and additionally take the ratio of the axes of the ellipse to measure how geographically directed or linear a social network is. The angle of the main axis represents the geographical orientation, interpreted as how culturally diverse the social network is. Using a Tobit model they found that young people with higher education and low to middle income, and those with more education or workplace moves, tend to maintain more spatially distributed networks. They also found that the spatial distribution of these education and workplace moves, measured as confidence ellipses, did not have significant influence on the spatial distribution of the social network. Kowald and Axhausen (24) employ a weighting scheme while calculating the confidence ellipses, weighting a contact's home location by the summed annual contact frequencies with the ego. Concerns have been raised by all of the aforementioned with the ellipse area as the size measure of a network in that there is an over-representation of space, partially due to the ellipses being symmetrical and the assumption of continuity.

While the confidence ellipse is easy to calculate and a useful measure of egocentric networks spatial distribution, there is room for improvement in travel behavior research. While the addition of the axes ratio and orientation add to the measure, the addition of travel time, spatial impedance, travel costs, and nonparametric distributions should be explored. In this study, we explore minimizing the number of contacts used and incorporating the contact to contact locations / distances.

3. Data

The People in Your Life (PiYL) pilot survey was designed to gauge the geographic extent of a respondents' social network and capture indicators of the level of their long-distance travel to facilitate modeling social network geography as a predictor of long-distance travel behavior. It was developed in response to focus group interviews in 2013 following a one-year panel tracking long-distance travel (25). The participants indicated that many of their long-distance travel choices where influenced by the location of family, friends, and work activities. Comprehensive documentation of all individuals' in a social network and long-distance travel behaviors are both highly burdensome tasks. Thus, a primary goal of the PiYL pilot was to test the effectiveness of collecting both more abbreviated social networks and more abbreviated travel data. After multiple rounds of testing and development, the pilot survey was administered in winter of 2016-2017. It collected home locations for only 13 individuals in each ego's social network, self-assessed travel frequency for eight different trip types, and a limited slate of demographic variables regarding the ego.

The pilot survey was administered using a paper survey to a total of 110 respondents recruited in Alabama, California, and Vermont. The Alabama-based respondents consisted of 65 engineering undergraduate and graduate students and several staff members at Auburn University and was handed back to the research team. Twenty-one California-based participants living in greater Sacramento were recruited at the University of California Davis or from senior citizen participants in a University seminar program, and twenty-four women were recruited from Burlington, Vermont through email and advertisements at social services organizations. California surveys were mailed back and Vermont surveys were handed back to the interviewer. Additional information about the creation of the PiYL pilot survey and the demographics of the respondents can be found in Aultman-Hall et al. (26).

The 13 contacts in each ego's social network consisted of 10 people defined based on their relationship to the ego (relation-based contacts) and three people selected based on home locations (location-based contacts). Respondents were asked to provide the home locations for 10 relation-based contacts according to the following criteria:

- three family members that did not live with the respondent;
- a person the respondent would go to for work or professional advice;
- a person the respondent would go to for personal advice;
- a good friend;
- a childhood friend;
- a person the respondent wishes they could spend more time with; and
- two people whom the respondent felt an obligation to visit.

Initial analysis has suggested average distance to the whole set of 19 relation-based contacts was highly correlated with average distance to the three family members (26). In addition, respondents were asked to identify contacts with whom they had communicated with in the last year that lived in specific, distant locations. The specified locations were large population states on the opposite side of the country. Europe and Asia were used for all respondents. The locations varied based on the respondent's home state and were selected based on discussions with pre-test respondents. These three contacts are referred to as location-based contacts. Contacts were solicited in the following locations:

- New York, California, and Europe/Asia for Alabama-based participants
- New York, Florida, and Europe/Asia for California-based participants
- Florida, California, and Europe/Asia for Vermont-based participants

General long-distance travel behavior measures were collected by asking the respondents to estimate the frequency with which they undertook the following eight non-exclusive trips types:

Trips to destinations more than a 2-hour drive from home:

- To visit family or friends;
- For work; and
- For personal business such as a medical appointment, banking, or other services.

Trips meeting the following criteria:

- For vacation or leisure;
- That include air travel;
- With NO overnight stay that include air travel;
- With NO overnight stay and include 2 or more hours of driving EACH way; and
- That include a destination outside of North America.

Trip frequencies were recorded on a six-point scale:

- More than once per Month,
- Once per Month,
- Multiple Times per Year,
- Once per Year,
- Less than Once per Year, and
- Never.

These long-distance travel measures are relatively broad and recent analysis of our prior one-year panel (25) and the PiYL travel frequency estimates (26) suggest that annual self-assessed travel frequency is not an accurate indicator of travel. Thus, these travel measures should be considered general travel levels and interpreted with caution. These travel metrics may be the reason for weak results associating average distance to social network to travel level (26). The social network and long-distance questions in the PiYL pilot survey were necessarily limited in their scope because a main goal of the survey pilot was to reduce survey burden. Our intention was to create a sub-set of contacts that could be representative of broader social network physical extent or geography. Collection of these pilot data allowed for development of preliminary types of social network geography which were tested for relationship to long-distance travel frequency.

While the PiYL survey participants were not recruited randomly, the 110 respondents did have significant variability in age, gender, education, and income (Table 1). The sample provided home locations for 992 relation-based contacts and 142 location-based contacts. Contacts were dispersed globally. Respondents tended to have fairly close emotional relationships with their relation-based contacts, reporting an emotional closeness of 7 or higher on an 11-point scale where 0 indicated "not at all close" and 10 indicated "very close" for 66% of the relational-contacts. Respondents were also in relatively frequent communication with these contacts, reporting face-to-face interaction within the last month with 51% of the contacts and telecommunications with 75% of the contacts in the same time frame. The physical distance between the respondents and their relational-contacts was highly variable and not significantly correlated with emotional closeness.

Table 1. Sample Description

Categorical Variables	Category	Total N	% of Total (N=110)
Gender	Male	58	52.7%
Employment Status (all binary variables)	Full-time	36	32.7%
	Part-time	32	29.0%
	Full-time Student	48	43.6%
	Not employed/Retired	13	11.7%
Household Size	4 or more	27	24.6%
	3	7	24.5%
	2	41	37.3%
	1	15	13.6%
Education	High School or Some HS	4	3.6%
	Some College	44	40.0%
	Bachelor's or Associate's	34	30.9%
	Graduate or Prof. Degree	28	25.5%
Cell Phone	Yes	109	99.1%
Income	$150,000 or more	15	13.7%
	$100,000-$149,999	20	18.2%
	$50,000-$99,999	19	17.3%
	$25,000-$49,000	17	15.5%
	< $25,000	24	21.9%
	Prefer not to answer	15	13.6%
Telecommute Type (all binary variables)	Yes, often	6	5.5%
	Yes, occasionally	26	23.6%
	No	69	62.8%

Age (years)	Mean
AL	25
CA	62
VT	43

Simple measures of social network extent, measured by the logged average distance to each respondent's contacts, was modeled against typical socio-demographic predictor variables: gender, age, presence of children in the household, household income, educational attainment, telecommuter status, and state of residence. The resulting models were weak but some variables, such as income were significant (26).

Travel frequency data collected for eight different trip types showed little internal correlation. The travel frequencies of different trip types were related. Thus, respondents who traveled frequently for one type of trip were not notably more likely to travel frequently for other types of trips. The respondents tended to show a similar overall level of long-distance travel across all trip types, though this finding may be an artifact of the small sample size and non-random sampling.

4. Social Network Results

A new social network classification method utilizing the distances from the ego to all of their contacts as well as the distances between all contact pairs in the respondent's social networks is presented here using the PiYL survey. Several conceptual scenarios, demonstrating different extremes in network pattern, were developed as an initial step towards classifying social networks. These six conceptual social network types are shown in Figure 1. In each plot, the ego is shown at the center as a black star and the contacts are distributed around them, indicated by colored dots. Images A and C in Figure 1 show two different extreme scenarios, where all of a person's closest contacts are at a far distance. In the first case (A) the contacts are distributed uniformly around the ego and in the second case (C) they are clustered in one direction with respect to the ego. Whether contacts are all in one direction or in many different directions might have an impact on travel by the ego, considering that if the contacts are all in one location then one trip could allow the ego to interact with all of their closest contacts.

Four distance-based measures were considered as the basis for the classifying networks quantitatively. These measures are the mean distance and variance in distance from the ego to each contact (referred to as the "ego-to-contact" or ETC measures) and the mean distance and variance in distance from each contact to every other contact, (referred to as the "contact-to-contact" or CTC measures). Conceptual ETC and CTC levels are also provided in Figure 1. While real-world social networks are expected to have more variability than those shown in these examples, different typologies of social network may be identifiable using cluster analysis of ETC and CTC measures. ETC and CTC distances of PiYL respondents were calculated using the latitudes and longitudes for each respondent's home city location and that of their contacts' city location using the great circle distance method. A summary of the ETC and CTC variables for the PiYL dataset can be found in Table 2.

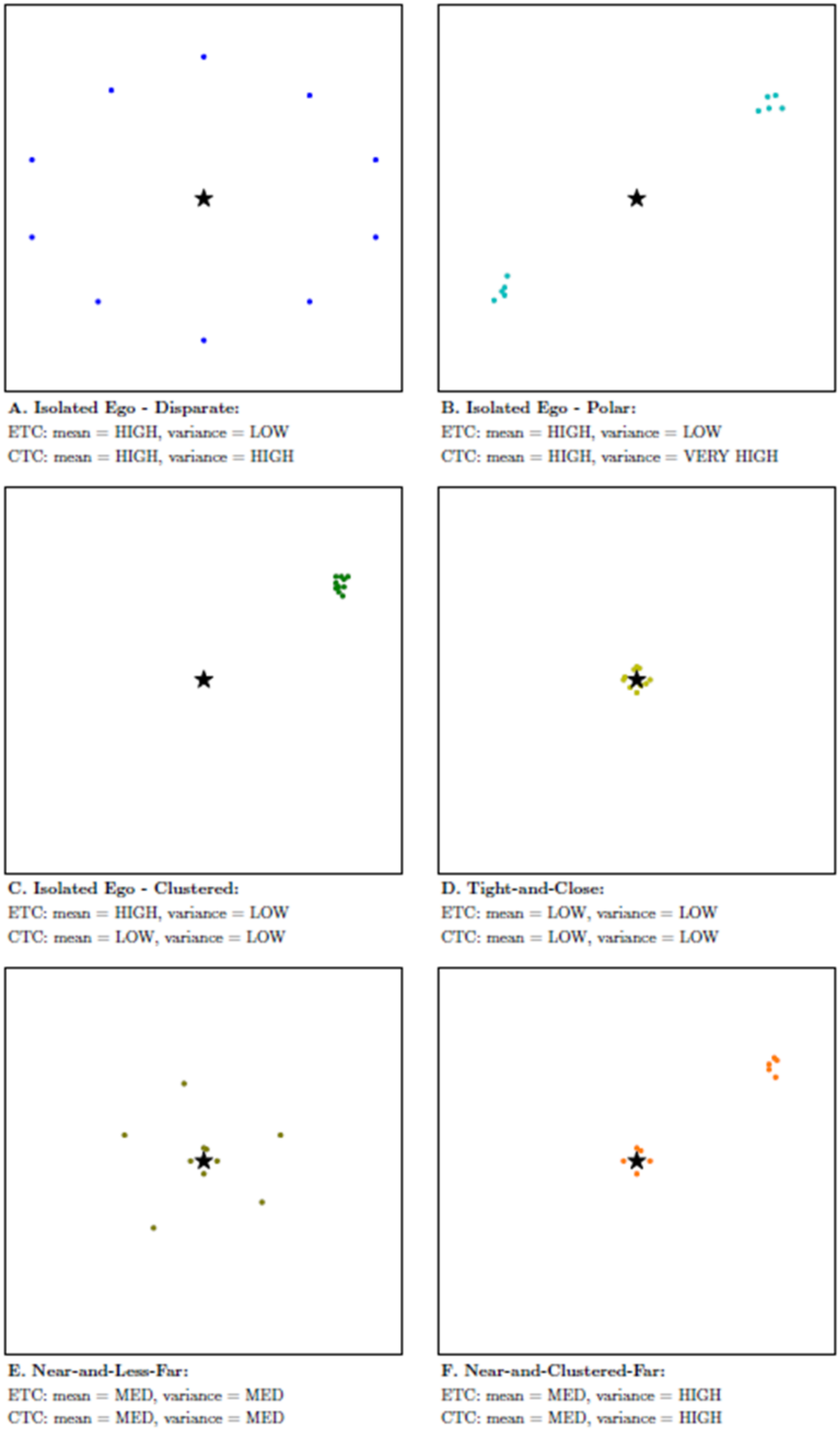

Figure 1. Conceptual Social Network Types

Table 2. Summary of Distance Variables for all Respondents' Social Networks

DISTANCE (MILES)		MEAN	STD. DEV.	MIN.	25%	50%	75%	MAX.
Ego-to-Contact	Average	523.5	699.3	24.2	121.6	329.6	619.1	5250.9
	Standard Deviation	585.7	663.7	21.9	149.9	355.6	782.1	4206.2
	Coefficient of Variance	24.0	13.4	4.5	13.4	20.5	33.7	73.4
Contact-to-Contact	Average	708.3	800.6	39.0	205.2	383.3	974.0	5006.6
	Standard Deviation	656.6	704.8	27.3	184.2	407.5	934.9	4084.7
	Coefficient of Variance	22.6	12.2	4.4	12.9	19.6	30.3	65.7
	Number of Contacts	9.1	1.2	5	8	10	10	10

K-Means clustering was performed on four candidate sets of ETC and CTC distance variables, shown in Table 3. The candidate sets compared the effectiveness of using the standard deviation to the coefficient of variance for the ETC/CTC as well as the inclusion/exclusion of a number of contacts as a clustering variable. The score distributions for these four candidate sets can be seen in Figure 2. Set 4 included the averages and coefficients of variance of the ETC and CTC distances for each respondent and was selected for the final clustering criteria because it achieved a higher score than clustering with the standard deviation. Inclusion or exclusion of the number of contacts variable had limited importance on cluster score and did not change how respondents were clustered.

Table 3. Candidate Clustering Variable Set

Variables	Variable Sets			
	Set 1	Set 2	Set 3	Set 4*
ETC Average Distance	X	X	X	X
ETC Coefficient of Variance		X	X	X
ETC Standard Deviation	X		X	
CTC Average Distance	X	X	X	X
CTC Coefficient of Variance		X	X	X
CTC Standard Deviation	X		X	
Number of Contacts	X	X		

* Final Variable Set

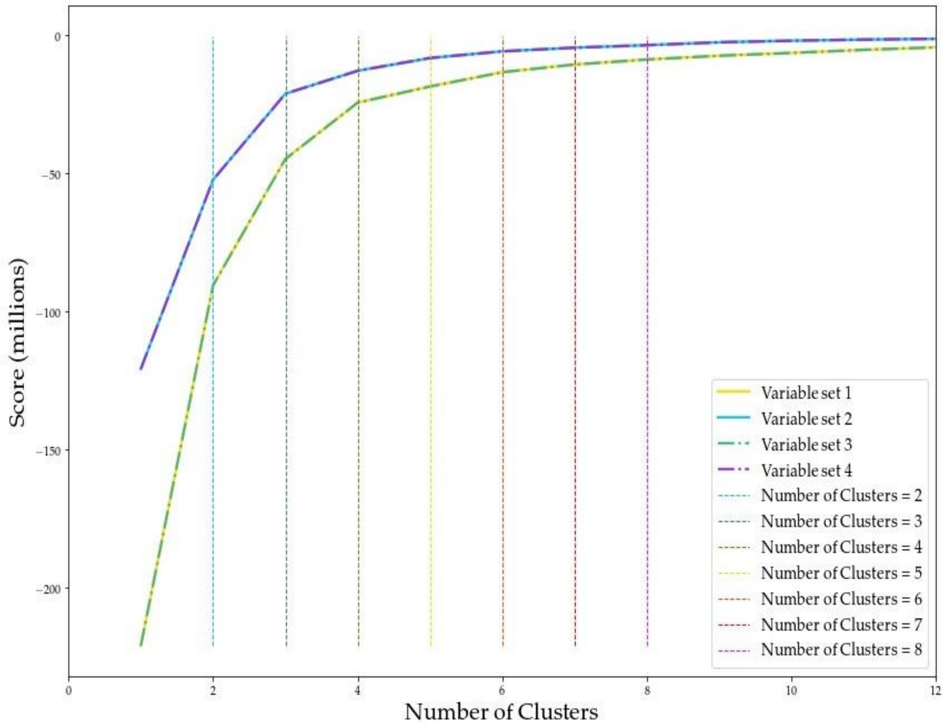

Figure 2. Scores of K-Means Cluster Candidate Variable Sets by Number of Clusters

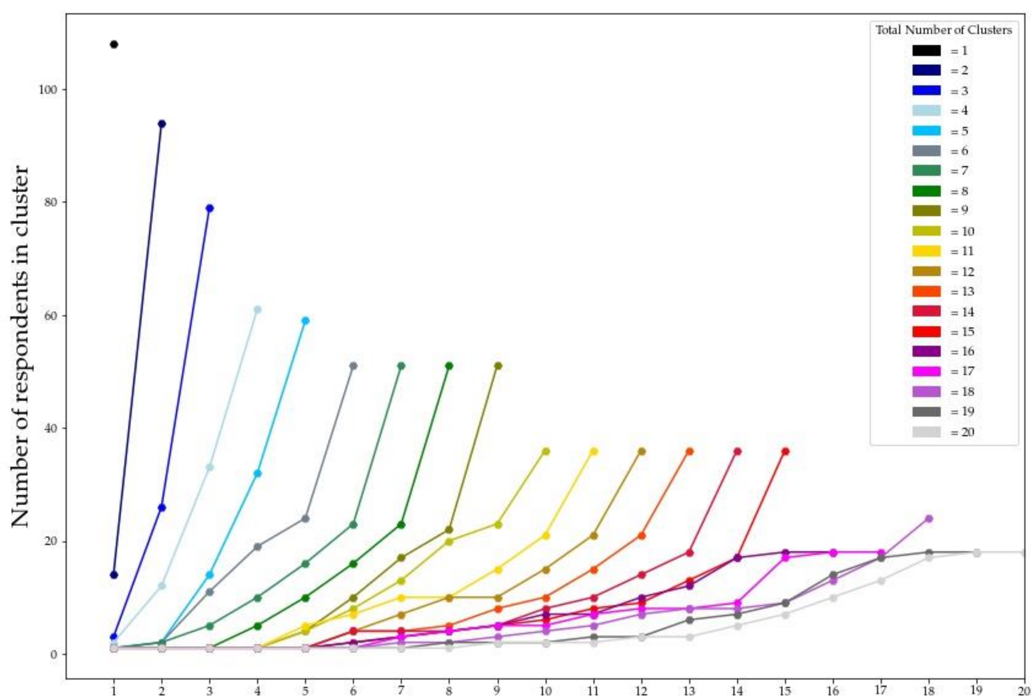

Figure 3. Clustering Distribution for Variable Set 4

The clustering distributions produced by clustering into 1 to 20 total clusters with the final clustering variable set is shown in Figure 3. The number of clusters used for the final analysis should result in meaningfully sized clusters – that is clusters that are small enough to distinguish among respondents based on important differences but not so small that minor differences between respondents separates them into different groups. At the most extreme scenarios, using a single cluster would group all respondents together while using as many clusters as respondents would result in each respondent having their own group – neither of which provide useful information about the respondents. Acknowledging the limitations of the very small sample size and based on the clustering distribution results, we assessed that a set of six clusters succeeded in creating unique groupings with this pilot data. These clusters reflect significant differences in terms of the social network geography variables. The additional groupings created when using more than six clusters were very small in size and not appreciably different from the groups produced with six clusters. For this reason, the final analysis was conducted with six clusters.

Once the PiYL respondents had been clustered into six groups, each respondent's social network was mapped and visually inspected. The six clusters were named based on the common characteristics of the spatial distributions of the social network. Categorizations incorporated the general distance from the ego to other contacts (regional, continental, or global) as well as the degree of concentration among the contacts (dispersed versus. polar – only two or three unique locations). The ETC and CTC variables for each of the six clusters are summarized in Table 4.

The largest group, Cluster 1, was characterized as "regional", since these social networks were dominated by contacts living in the same region as the ego. The 51 respondents whose social network geographies were regional had fairly low average long-distance trip frequencies using air and to international destinations, but the highest average frequency for visiting family and friends. This group was 66% male and dominated (75%) by respondents between the ages of 21 and 24 years old. The regional cluster was also proportionally lower income and less educated than the full sample. Recall these respondents may be the Auburn students. The second largest cluster, Cluster 6, consisted of 24 respondents with "polar continental" social networks, meaning they were contained within the country or continent of the ego, and that most contacts lived very close to the ego while a small number lived very far away in only one or two unique locations. This cluster had fairly high average long-distance trip frequencies in general, though not for international travel. It was 58% female and had an average age of 35 years. Cluster 3 was categorized by "dispersed global" social networks and included 11 respondents. This cluster was predominantly male, highly educated (64 % had Graduate or Professional Degrees), older, and had high long-distance travel frequencies for all but one of the respondents. As we might assume, this cluster also was proportionally higher income, with no respondents reporting less than $25,000, 45.5% reporting household incomes greater than $100,000 but 27.3% not reporting at all. The last of the larger clusters is Cluster 4, which was predominantly "dispersed continental" networks. These respondents were 70% female, had the oldest average age, 50 years old, and had very similar average long-distance trip frequencies to Cluster 6, the other "continental" cluster. Clusters 2 and 5 contained only three individuals in total and consisted of networks where contacts' homes were on the opposite side of the world from the ego. These clusters were uncommon in this small sample, possibly due to sample size, and need to be assessed further, both in terms of our clustering techniques and our calculation of distances when the contacts are halfway around the world.

Table 4. Summary of Cluster Variables by Cluster Type

Cluster Type	Distance (Miles)	Mean	Std	Min	Max
Cluster 1: Regional n=51	ETC Average	125.8	59.6	24.2	272.7
	ETC Coefficient of Variance	15.1	6.6	4.5	37.7
	CTC Average	201.4	97.4	39.0	377.7
	CTC Coefficient of Variance	14.3	6.2	4.4	35.7
Cluster 2: Polar Global n=2	ETC Average	2682.6	597.6	2260.0	3105.2
	ETC Coefficient of Variance	57.6	2.3	56.0	59.2
	CTC Average	3783.1	748.6	3253.7	4312.5
	CTC Coefficient of Variance	50.3	1.2	49.4	51.1
Cluster 3: Dispersed Global n=11	ETC Average	1411.9	372.4	1026.9	2380.9
	ETC Coefficient of Variance	39.6	16.0	19.4	73.4
	CTC Average	1752.9	352.1	1312.7	2376.5
	CTC Coefficient of Variance	38.2	13.0	23.8	65.7
Cluster 4: Dispersed Continental n=19	ETC Average	746.3	183.1	516.5	1192.3
	ETC Coefficient of Variance	31.2	7.1	19.9	42.6
	CTC Average	1073.2	144.4	827.2	1353.7
	CTC Coefficient of Variance	27.5	6.2	19.5	42.2
Cluster 5: Polar Global n=1	ETC Average	5250.9	N/A	5250.9	5250.9
	ETC Coefficient of Variance	58.0	N/A	58.0	58.0
	CTC Average	5006.6	N/A	5006.6	5006.6
	CTC Coefficient of Variance	57.7	N/A	57.7	57.7
Cluster 6: Polar Continental n=24	ETC Average	408.1	71.6	262.7	583.4
	ETC Coefficient of Variance	25.9	10.0	8.5	41.6
	CTC Average	582.5	153.1	298.5	889.1
	CTC Coefficient of Variance	25.3	9.4	10.3	39.9

Given the pilot nature of the work and small sample size we are not seeking to generalize any of these results to the broader population. However, a breakdown of travel frequency for the four larger social network clusters is presented in Table 5 for illustration. As expected, trips involving air travel and to non-North American destinations (note that these are overlapping categories) were more common among respondents with continental and global social networks than those with regional clusters. Conversely, respondents with regional social networks had the higher frequency of visiting family and friends. Both of these results are consistent with the hypothesis that social network extent influences personal travel decision-making.

Table 5. Travel Frequency by Social Network Cluster Type

Trip Type	Trip Frequency	SOCIAL NETWORK CLASSIFICATION CLUSTERS			
		Regional	Polar Continental	Dispersed Continental	Dispersed Global
Trips to Visit Family/Friends[1]	once per month or more	45%	25%	11%	27%
	multiple times per year	43%	63%	74%	45%
	once a year or less	10%	13%	16%	18%
	never	2%	0%	0%	9%
Personal Business Trips[1]	once per month or more	8%	8%	0%	9%
	multiple times per year	22%	13%	0%	0%
	once a year or less	29%	38%	26%	36%
	never	39%	42%	68%	45%
Work Trips[1]	once per month or more	10%	13%	0%	27%
	multiple times per year	20%	17%	26%	18%
	once a year or less	25%	25%	42%	18%
	never	43%	46%	26%	27%
Vacation or Leisure Trips	once per month or more	10%	17%	5%	18%
	multiple times per year	65%	54%	79%	55%
	once a year or less	24%	29%	16%	27%
	never	2%	0%	0%	0%
Trips Including Air Travel	once per month or more	0%	0%	0%	9%
	multiple times per year	14%	54%	47%	36%
	once a year or less	65%	46%	47%	55%
	never	22%	0%	0%	0%
Air Trips with No Overnight Stay	once per month or more	0%	0%	0%	0%
	multiple times per year	4%	0%	0%	9%
	once a year or less	12%	25%	26%	18%
	never	82%	75%	63%	73%
Driving Trips With No Overnight[2]	once per month or more	6%	13%	5%	9%
	multiple times per year	33%	33%	21%	18%
	once a year or less	33%	29%	47%	45%
	never	27%	25%	21%	27%
Trips Out of North America	once per month or more	0%	0%	0%	0%
	multiple times per year	0%	0%	0%	18%
	once a year or less	39%	67%	79%	64%
	never	61%	33%	16%	18%
Total respondents in cluster		51	24	19	11

[1] Trips to a destination more than 2 hours from where the respondent currently lives.
[2] Including 2 or more hours of driving each way.

5. Conclusions

The findings based on this pilot data collection result in three basic conclusions. First, the results agree with prior studies that social network extent is related to types and levels of long-distance travel. Collection of comprehensive long-distance travel data is equally burdensome to social network data collection yet the small datasets and studies conducted over the last decade, including this one, support the potential of social network attributes being a valuable predictor of travel. There is a logical conjecture that if a significant portion of long-distance travel is personal or leisure then there is an interrelated causal relationship that a wide social network leads to more travel. Moreover, participating in long-distance travel supports and possibly extends one's social network.

Second, the new approach presented here categorizing social networks using not only the distances from the ego to their contacts, but also the distances between each contact in the social network is able to incorporate the geographic extent and shape of the networks when compared to the more basic approach (e.g. the average distance to contact method). Moreover, reasonable clusters were created using this small sample. Preliminary examination of the small PiYL dataset shows coherent patterns in the estimated travel behavior for the respondents in the larger clusters. Full development of representative clusters will require a larger dataset.

Third, the research here was not conducted on a full or comprehensive social network suggesting that there is merit to the idea that a representative, but not comprehensive, social networks may be adequate for transportation-related research. If future research could comprehensively validate this proposition, the burden of adding social network measures to travel surveys would be reduced and potentially manageable.

Further work should be conducted with a larger sample size to analyze this method of categorization of social network geography against other continuous methods such as confidence ellipse area. As discussed above, future research should also use a different, more accurate measure of level of long-distance travel potentially derived from passive or semi-passive mobile devices to eliminate the inaccuracy of recall for longer study periods and also to reduce participant burden.

References

1. Axhausen, K. Personal biography, social networks and travel behaviour: Hypotheses and assumptions. In Presentation at Odyssey Meeting, University of Ulster, Belfast, August (2004).

2. Axhausen, K. Geographies of social networks. In 37th World Congress of the International Institute of Sociology (2005), Frontiers of Sociology.

3. Axhausen, K. W. Social networks and travel: Some hypotheses. Social dimensions of sustainable transport: transatlantic perspectives (2005), 90–108.

4. Axhausen, K. W. Activity spaces, biographies, social networks and their welfare gains and externalities: some hypotheses and empirical results. Mobilities 2, 1 (2007), 15–36.

5. Larsen, J., Axhausen, K. W., and Urry, J. Geographies of social networks: meetings, travel and communications. Mobilities 1, 2 (2006), 261–283.

6. Larsen, J., Urry, J., and Axhausen, K. Mobilities, Networks, Geographies, paperback 2016 ed. Routledge, 2006.

7. Wellman, B. Network analysis: Some basic principles. Sociological theory (1983), 155–200.

8. Marsden, P. V. Network data and measurement. Annual review of sociology 16, 1 (1990), 435–463.

9. Burt, R. S. Network items and the general social survey. Social networks 6, 4 (1984), 293–339.

10. Campbell, K. E., and Lee, B. A. Name generators in surveys of personal networks. Social networks 13, 3 (1991), 203–221.

11. Carrasco, J. A., Bustos, C., and Cid-Aguayo, B. Affective personal networks versus daily contacts: Analyzing different name generators in a social activity-travel behavior context. In Transport survey methods: best practice for decision making. Emerald Group Publishing Limited, 2013, pp. 409–426.

12. Marin, A., and Hampton, K. N. Simplifying the personal network name generator: Alternatives to traditional multiple and single name generators. Field methods 19, 2 (2007), 163–193.

13. Kim, J., Rasouli, S., and Timmermans, H. J. Social networks, social influence and activity-travel behaviour: a review of models and empirical evidence. Transport Reviews (2017), 1–25.

14. Huang, G., and Tausig, M. Network range in personal networks. Social Networks 12, 3 (1990), 261–268.

15. Killworth, P. D., Johnsen, E. C., Bernard, H. R., Shelley, G. A., and McCarty, C. Estimating the size of personal networks. Social Networks 12, 4 (1990), 289–312.

16. Kowald, M., van den Berg, P., Frei, A., Carrasco, J.-A., Arentze, T., Axhausen, K., Mok, D., Timmermans, H., and Wellman, B. Distance patterns of personal networks in four countries: a comparative study. Journal of Transport Geography 31 (2013), 236–248.

17. Carrasco, J., Miller, E., and Wellman, B. How far and with whom do people socialize?: Empirical evidence about distance between social network members. Transportation Research Record: Journal of the Transportation Research Board, 2076 (2008), 114–122.

18. Frei, A., and Axhausen, K. W. Modeling spatial embedded social networks. Working paper Transport and Spatial Planning 685 (2011).

19. Axhausen, K. W., and Frei, A. Size and structure of social network geographies. Arbeitsbericht Verkehrs-und Raumplanung 444 (2007).

20. Schönfelder, S., and Axhausen, K. W. Activity spaces: measures of social exclusion? Transport policy 10, 4 (2003), 273–286.

21. Schönfelder, S. Urban rythms: modelling the rythms of individual travel behaviour. department bau. Umwelt und Geomatik, ETH, Zu ̈rich (2006).

22. Schönfelder, S., and Axhausen, K. W. Structure and innovation of human activity spaces. Arbeitsberichte Verkehrs-und Raumplanung 258 (2004), 1–40.

23. Guidon, S., Wicki, M., Axhausen, K. W., and Bernauer, T. Investigating the relationship between individuals' social networks, mobility behaviour, and social capital-survey method and first results. In 17th Swiss Transport Research Conference (STRC 2017) (2017), STRC.

24. Kowald, M., and Axhausen, K. W. The structure and spatial spread of egocentric leisure networks. In 7th International Conference on Applications of Social Network Analysis, ASNA 2010. Abstracts (2010), vol. 651, ETH, pp. 65–66.

25. Aultman-Hall, L., Harvey, C., LaMondia, J.J., and Ritter, C. Design and response quality in a one-year longitudinal survey of overnight and long-distance travel. Transportation Research Procedia 11 (2015), 136-153

26. Aultman-Hall, Lisa, Jeffrey J. LaMondia, Hannah Ullman, and Madeline Suender. (2017) Surveying Social Network Geography to Model Long-distance Travel. International Conference on Travel Survey Methods, L'Estrel Quebec.